JOINT CLASSIFICATION AND TRAJECTORY REGRESSION OF ONLINE HANDWRITING USING A MULTI-TASK LEARNING APPROACH

Felix Ott[1,2], David Rügamer[2], Lucas Heublein[1], Bernd Bischl[2] and Christopher Mutschler[1]

[1] Fraunhofer IIS, Fraunhofer Institute for Integrated Circuits IIS; {felix.ott | heublels | christopher.mutschler}@iis.fraunhofer.de
[2] LMU Munich, Munich, Germany; {david.ruegamer | bernd.bischl}@stat.uni-muenchen.de

MOTIVATION

Idea:

- Handwriting is important for the *graphomotoric*, i.e., the learning process of students. State-of-the-art methods to recognize handwriting (a) require to write on a special device (affecting the writing style), (b) require to take images of the text (OCR), or (c) are only prototypical.
- The sensor-enhanced pen by STABILO International GmbH allows to write on normal paper that we use for trajectory reconstruction and text classification.

Problem:

- Reconstructing the pen tip trajectory with distance-based losses does not provide smoothed trajectories.
- Maximizing the similarity is scale invariant.

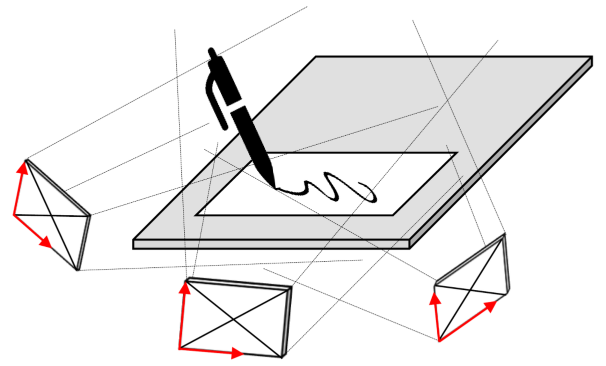
Goal:

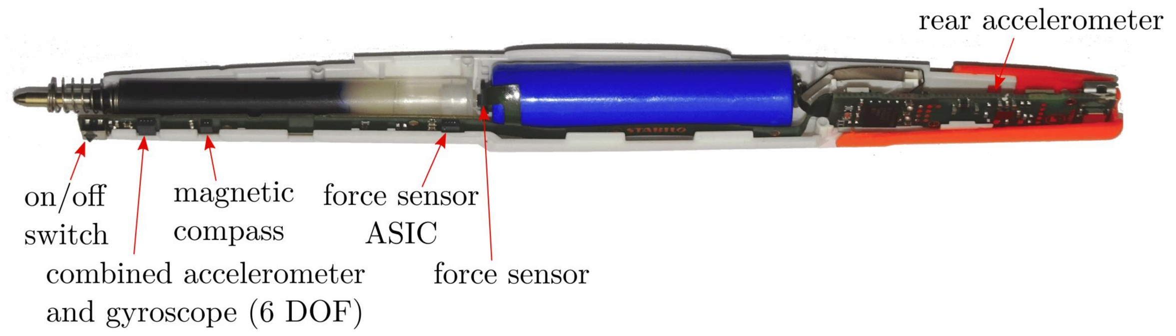
- Combine *distance*, *spatio-temporal* and *distribution* losses with Multi-Task Learning techniques.
- Combine classification and trajectory reconstruction task.

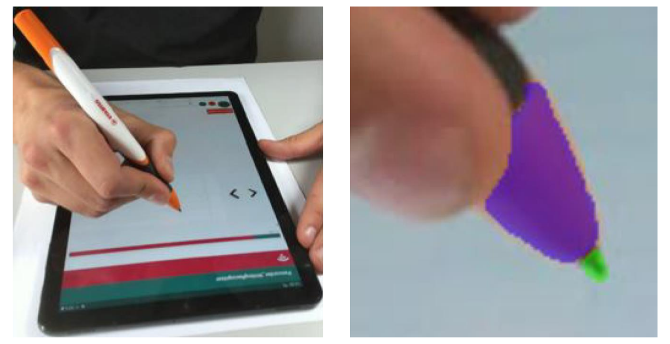
Input	Task	Loss	Output
Inertial or visual MTS	Classification	Cross-entropy	Character class or/and trajectory (MTS)
	Regression	Distance, Spatio-temporal, Distribution	

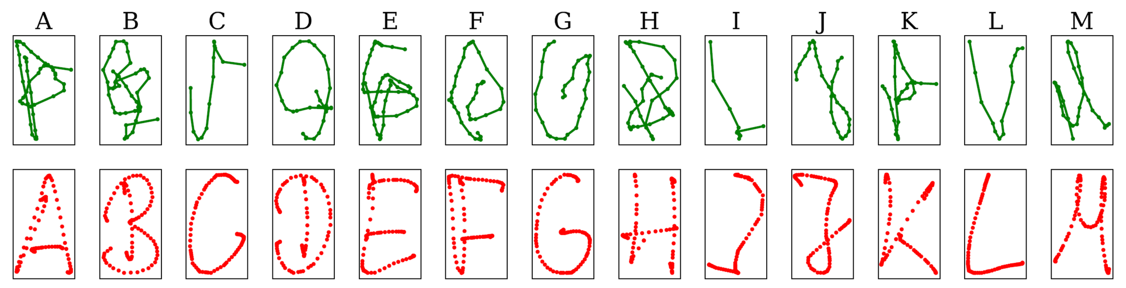
RECORDING SETUP

- Sensor-enhanced pen with two accelerometers, a gyroscope, a magnetometer and a force sensor (at 100 *Hz*).
- Writing on a tablet for ground truth trajectory (at 30 *Hz*).
- Three outside-in cameras for pen tip tracking segmented with U-Net (at 60 *Hz*).
- Total 2,466 training and 992 validation characters.
- Dataset publicly available: https://iis.fraunhofer.de/onhw-dataset/









METHODOLOGY

Trajectory Reconstruction Loss Functions:

- Distance-based losses: mean squared error $\mathcal{L}_{MSE}$, Huber loss $\mathcal{L}_{H}$, Andrew's Sine loss $\mathcal{L}_{AS}$
- Spatio-temporal losses: Cosine Similarity $\mathcal{L}_{CS}$, Pearson Correlation $\mathcal{L}_{PC}$
- Distribution-based loss: Wasserstein distance $\mathcal{L}_{WAS_p}$

Architectures With Different Split Points:

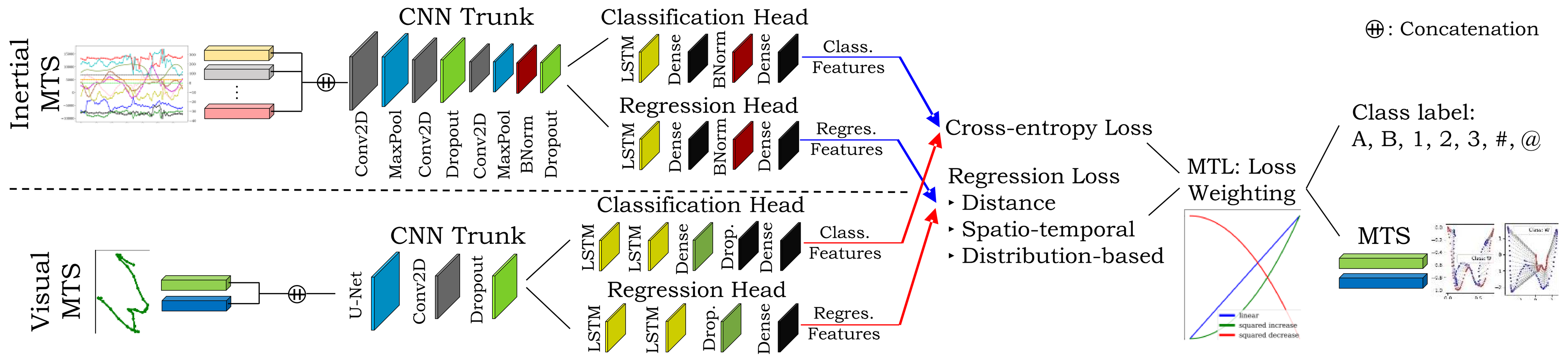
- Baselines: A_0 only regression, A_1 only classification
- A_2 classification for regression, A_3 latest split, A_4 late split, A_5 to A_7 intermediate splits, A_8 separate heads

Multi-Task Learning Approaches:

- Naive and epoch-dependent weighting of the second task.
- Dynamic Weight Average (DWA) by averaging task weighting over time: weights for current epoch e are

$$\omega_i(e) = \frac{e^{\lambda_i(e-1)/P}}{\sum_k e^{\lambda_k(e-1)/P}};\ \lambda_i(e-1) = \frac{\mathcal{L}_i(e-1)}{\mathcal{L}_i(e-2)}$$

with P as pre-specified softness of task weighting.



Method overview: multivariate time series input (top: inertial data, bottom: visual data). Classification (*cross-entropy*) and regression (*distance*, *spatio-temporal* and *distribution*-based) loss functions are combined with different MTL weighting strategies.

EXPERIMENTAL RESULTS

MTL Architecture Evaluation:

- Baseline: Inertial: 0.1705 rmse / 88.11%
  Visual: 0.1360 rmse / 73.19%
- Inertial: improves up to 0.1169 (model A_2) to 0.1623 (model A_6)
- A late split has a positive influene on the trajectory regression by sharing more trainable parameters in the trunk.

Loss Function Evaluation:

- $\mathcal{L}_H$ increases the classification accuracy up to 88.43%.
- Spatio-temporal losses are able to learn the shape of the character, but at a wrong scale. $\mathcal{L}_{PC}$ is smoother compared to $\mathcal{L}_{CS}$.

Combined Loss Function Evaluation:

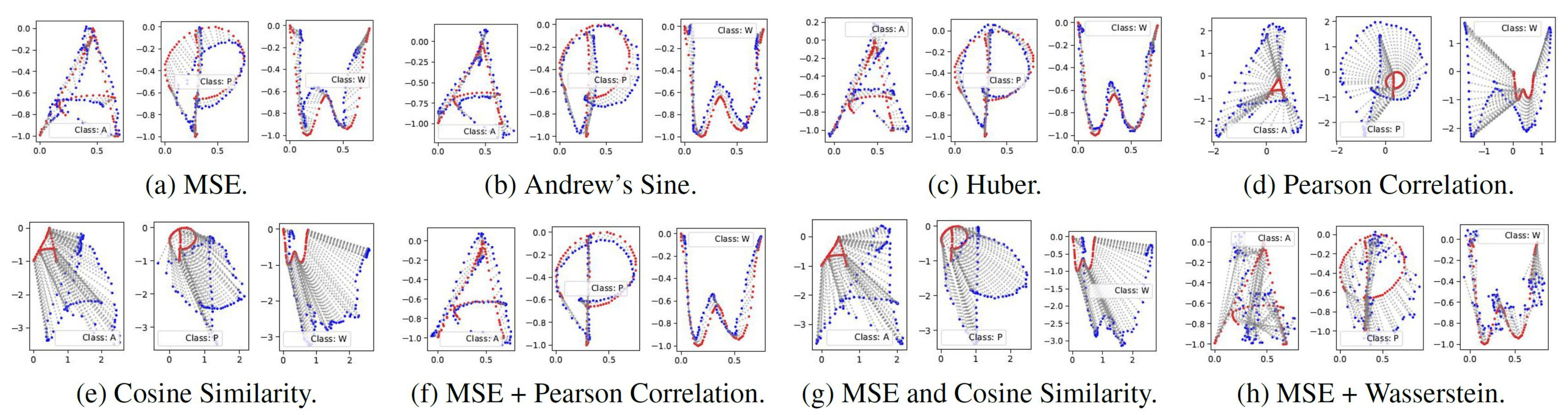
- $\mathcal{L}_{MSE}$ combined with $\mathcal{L}_{WAS_p}$: the predictions become more evenly distributed for larger p. For $p=1$ and the naive weighting strategy, the trajectory prediction error is large, but can be significantly improved using alternative MTL strategies.
- $\mathcal{L}_{MSE}$ combined with $\mathcal{L}_{PC}$ provides a suitable shape while still yielding to an improvement in accuracy. DWA is worse than linear or squared weight increase.
- The order of weight increase is important for combining metrics. A slow weight increase is the best approach for jointly learning the classification and trajectory regression tasks.

Network	MSE+CE		AS+CE		H+CE		MSE+PC+CE		MSE+CS+CE		MSE+WAS+CE		PC+CE	CS+CE	WAS+CE
	Traj.	Class.	Traj.	Class.	Traj.	Class.	Traj.	Class.	Traj.	Class.	Traj.	Class.	Class.	Class.	Class.
Only regression (A_0)	0.1705	-	**0.1594**	-	0.1501	-	0.1723	-	1.0023	-	0.3107	-	-	-	-
Only classification (A_1)	-	88.11	-	-	-	-	-	-	-	-	-	-	-	-	-
Class. for regr. (A_2)	0.1169	86.69	0.1779	9.78	**0.1290**	62.78	**0.1127**	86.81	0.3554	86.28	0.1612	7.28	86.73	86.02	12.60
Latest split (A_3)	0.1381	86.67	0.1856	49.31	0.1569	66.73	**0.1381**	85.75	6.1464	87.22	0.3375	20.79	86.22	84.15	25.53
Late split (A_4)	0.1372	86.46	0.1421	76.28	0.1581	63.64	**0.1357**	**88.64**	1.3928	87.62	0.3262	26.65	86.67	**89.51**	29.74
Split after LSTM (A_5)	0.1370	87.34	0.1629	68.64	0.1458	73.74	**0.1386**	85.53	1.0578	**88.58**	0.3284	35.49	86.93	88.03	54.84
Split after 2. Drop. (A_6)	0.1623	87.68	0.1464	83.96	0.1580	84.76	**0.1647**	84.94	1.0053	85.28	0.3208	80.59	83.37	84.49	84.65
Split after 1. Drop. (A_7)	0.1866	84.27	0.1676	**86.93**	0.1546	86.89	**0.1638**	84.55	1.1388	87.20	0.3071	84.13	83.58	86.34	81.87
Separate heads (A_8)	0.1936	86.87	0.1660	86.02	**0.1533**	**88.43**	**0.1490**	87.03	1.0986	85.79	0.3315	82.03	87.15	**88.15**	82.58

Evaluation results for the IMU-based dataset trained with different loss combinations. Metrics: RMSE and accuracy (%).

Network	MSE+CE		AS+CE		H+CE		MSE+PC+CE		MSE+CS+CE		MSE+WAS+CE		PC+CE	CS+CE	WAS+CE
	Traj.	Class.	Traj.	Class.	Traj.	Class.	Traj.	Class.	Traj.	Class.	Traj.	Class.	Class.	Class.	Class.
Only regression (A_0)	0.1360	-	0.1388	-	0.1271	-	0.1348	-	2.5733	-	0.2302	-	-	-	-
Only classification (A_1)	-	73.19	-	-	-	-	-	-	-	-	-	-	-	-	-
Class. for regr. (A_2)	**0.1250**	**80.15**	0.1279	12.47	0.1475	55.81	0.1327	77.27	0.4900	74.33	0.1844	56.44	74.85	73.52	71.10
Latest split (A_3)	0.4314	23.21	0.1327	74.3	0.1577	64.15	**0.1401**	**75.54**	0.6713	9.32	0.1960	49.22	76.57	10.34	75.08
Late split (A_4)	0.1351	78.49	0.1260	78.23	**0.1230**	79.42	**0.1284**	**80.89**	1.1177	77.08	0.1348	74.50	58.07	74.42	74.25
Split after LSTM (A_5)	0.1291	**80.40**	**0.1252**	**81.24**	0.1307	78.06	0.1478	73.21	0.1705	77.22	0.1359	76.64	75.17	79.21	76.47
LSTM in tr. head (A_6)	0.1334	77.33	0.1383	74.31	0.1605	63.31	**0.1243**	**81.64**	0.1747	78.80	0.1376	73.64	74.05	77.47	75.42
LSTM in cl. head (A_7)	0.1378	78.58	0.1267	80.52	0.1330	78.86	0.1459	74.25	1.2483	79.99	0.1380	75.87	**80.51**	72.10	78.03
Split after 1. Drop. (A_8)	**0.1246**	78.56	0.1248	75.69	0.1310	75.18	0.1251	**79.27**	0.4132	78.67	0.2448	75.83	78.83	**80.76**	**79.42**

Evaluation results for the visual-based dataset trained with different loss combinations. Metrics: RMSE and accuracy (%).



(a) MSE. (b) Andrew's Sine. (c) Huber. (d) Pearson Correlation.
(e) Cosine Similarity. (f) MSE + Pearson Correlation. (g) MSE and Cosine Similarity. (h) MSE + Wasserstein.

Trajectory prediction (blue) agains the ground truth trajectory (red) of the characters ´A´, ´P´ and ´W´ based on inertial data.

Fraunhofer IIS
LMU Ludwig-Maximilians-Universität München
FUNDED BY Bundesministerium für Bildung und Forschung